CLINICAL INVESTIGATION PLAN

Artificial intelligence (AI) enabled decision support tool for selection of patients for lumbar spine surgery

Version number: 3.0

Date: 2024.11.05

Sponsor: University Hospital of North Norway
(Universitetssykehuset Nord-Norge HF)
Hansine Hansens veg 67
9019 TROMSØ

Coordinating investigator Tor Ingebrigtsen
Professor and consultant neurosurgeon
University Hospital of North Norway, Tromsø, Norway
Email: tor.ingebrigtsen@unn.no.

Revision history

Document version	Date of Issue	Summary of Change
1.0	2024.03.22	
1.1	2024.05.03	
3.0	2024.11.05	Revision after feedback from REK KULMU, the monitor and UNN's Data protection officer

Signatures

Sponsors representative
I am responsible for ensuring that this clinical investigation plan (CIP) includes all essential information to be able to conduct this clinical investigation. It is also my responsibility to appoint a dedicated independent monitor and approve the monitoring plan.

I will submit the CIP and all other important investigation-related information to the responsible investigator(s) so that they can conduct the clinical correctly. I am aware that it is my responsibility to hold the staff members who work with this clinical investigation informed and trained.

Sponsor's signature

_____

Tove Skjelbakken, director of research and education Date

Coordinating investigator
I have read this CIP and agree that it includes all essential information to be able to conduct the clinical investigation. By signing my name below, I agree to conduct the clinical investigation in compliance with this CIP, the Declaration of Helsinki, SS-EN ISO14155:2020 (Good Clinical Practice), and the current national and international regulations governing the conduct of this clinical investigation.

I will submit this CIP and all other important investigation-related information to the staff members and investigators who participate in this clinical investigation, so that they can conduct the clinical investigation correctly. I am aware of my responsibility to continuously keep the staff members and investigators informed and trained.

I am aware that quality control of this clinical investigation will be performed in the form of monitoring, audit, and possibly inspection.

Coordinating investigator's signature

_____

Tor Ingebrigtsen, professor and consultant neurosurgeon Date

Principal investigators

I have read this CIP and agree that it includes all essential information to be able to conduct the clinical investigation. By signing my name below, I agree to conduct the clinical investigation in compliance with this Clinical investigation plan, the Declaration of Helsinki, SS-EN ISO14155:2020 (Good Clinical Practice), and the current national and international regulations governing the conduct of this clinical investigation.

I will submit this CIP and all other important clinical investigation-related information to the staff members who participate in this clinical investigation, so that they can conduct the clinical investigation correctly. I am aware of my responsibility to continuously keep the staff members who work with this clinical investigation informed and trained.

I am aware that quality control of this clinical investigation will be performed in the form of monitoring, audit, and possibly inspection.

Principal investigator's signature

Tor Ingebrigtsen, professor and consultant neurosurgeon Date

Principal investigator's signature

Cato Kjærvik, associate professor and consultant orthopedic surgeon Date

Contact information

Roles	
Sponsors representative	Tove Skjelbakken Director of research and education University hospital of North Norway, Tromsø. tove.skjelbakken@unn.no +47 930 48973
Monitor	Ann Helen Jakobsen Research monitor kliniskmonitor@unn.no +47 951 31862
Coordinating and Principal investigator • Main responsibility for planning and conduction of the study, including analysis and publication • Responsible for recruitment and follow-up of surgeons at the University Hospital of North Norway • Main supervisor for the ph.d.-candidate • Emergency contact	Tor Ingebrigtsen Professor and consultant neurosurgeon University Hospital of North Norway, Tromsø tor.ingebrigtsen@unn.no +47 911 99843
Sub-investigator • Co-responsibility for planning and conduction of the study, including analysis and publication • Co-supervisor for the ph.d.-candidate	Tore Solberg Professor and consultant neurosurgeon, University Hospital of North Norway, Tromsø tore.solberg@unn.no +47 913 64531
Sub-investigator • Co-responsibility for planning and conduction of the study, including analysis and publication • Co-supervisor for the ph.d.-candidate	Karl Øyvind Mikalsen Associate professor and center director, The Norwegian center for clinical artificial intelligence, University Hospital of North Norway, Tromsø karl.oyvind.mikalsen@unn.no +47 915 65867
Principal investigator • Responsibility for recruitment and follow-up of surgeons and conduction of the study at the Nordland Hospital	Cato Kjærvik Associate professor and consultant orthopedic surgeon Nordland hospital, Bodø cato.kjaervik@nordlandssykehuset.no

• Participant in analysis and publication	+47 416 62667
Sub-investigator and ph.d.-candidate • Responsible for recruitment and follow-up of patients at the University Hospital of North Norway • Responsible for analysis of data and publication (under supervision)	Henrik Lykke Joakimsen hjo121@uit.no Phone +47 934 29542
Sub-investigator • Responsible for recruitment and follow-up of patients at the Nordland Hospital • Participant in analysis of data and publication	Oliver Muller Consultant orthopedic surgeon omu001.Mueller@nordlandssykehuset.no +47 476 46183.
IT-implementer • Participant (voluntary) in analysis of data and publication	Department for e-health, integrated care and innovation University Hospital of North Norway, Tromsø Participants: Eli Larsen eli.larsen@unn.no
IT-implementer • Participants (voluntary) in analysis of data and publication	Northern Norway regional health authority clinical IT-implementation project organization Participants: Heidi Johansen Project manager Northern Norway Regional Health Authority, Tromsø heidi.johansen@helsenord.no
Data processor and IT-developer • Participants (voluntary) in analysis of data and publication	Northern Norway regional health authority ICT trust Participants: Arne Magnus Sandvik, product owner Torkil Grindstein, quality registry consultant Are Edvardsen, innovation consultant Northern Norway Regional Health Authority ICT Trust, Tromsø arne.magnus.sandvik@hn-ikt.no

Industry partner (manufacturer) • Participants (voluntary) in analysis of data and publication	DIPS ASA Participants: Liv Bollvåg, research director Jørgen Aarmo Lund, ph.d.-student Thor Stenbæk, senior system architect Bjørn Næss, software developer lgb@dips.no
Industry partner • Participant (voluntary) in analysis of data and publication	Deepinsight AS Participants: Olav Willumsen Haugå, CEO Email: olav@deepinsight.no

Funding and research agreement

Funding for the studies is provided by the Northern Norway regional health authority (innovation grant), UiT The Arctic University of Norway (ph.d.-grant) and the Norwegian Medical Association (quality improvement grant). All project partners contribute in-kind.

List of acronyms and abbreviations

Abbreviation	Term/Explanation
AE	Adverse event
AI	Artificial intelligence
CIP	Clinical investigation plan
AUC	Area under the curve
DST	Decision support tool, interchangeably referred to as “device”
EHR	Electronic health record
EQ-5D-5L	EuroQol five-dimensions five-levels (measure of health-related quality of life)
GPE	Global perceived effect
HN IKT	Northern Norway regional health authority information technology trust (Helse Nord IKT)
HN FRESK	Northern Norway regional health authority clinical IT-implementation project organization (Helse Nord FRESK)
IT	Information technology
LDH	Lumbar disc herniation
LSS	Lumbar spinal stenosis
MDR	Medical devices regulation
ML	Machine learning
NORspine	The Norwegian registry for spine surgery
NRS	Numeric rating scale

ODI	Oswestry disability index
PASS	Patient acceptable symptom state
PROM	Patient reported outcome measure
REC	Regional committee for medical and health sciences research ethics
ROC	Receiver operated characteristic
SAE	Serious adverse event
SDM	Shared decision making
SPKI	The Norwegian center for clinical artificial intelligence (Senter for pasientnær kunstig intelligens)
UNN	University hospital of North Norway
UiT	UiT The arctic university of Norway
WP	Work package

1. Synopsis

English

Background
One third of patients operated for lumbar disc herniation (LDH) or spinal stenosis (LSS) do not achieve substantial improvement. Studies indicate that well informed shared decision making (SDM) can improve the selection to surgery, and thus the outcomes. Numerous algorithms for outcome prediction have therefore been developed, and some use artificial intelligence (AI). Most are trained on small datasets, few are accurate, all are stand-alone or web-based applications not integrated in the electronic health record (EHR), and none are implemented in routine clinical practice.

The Norwegian registry for spine surgery (NORspine) comprises a cohort of more than 69,000 cases. We have used AI to analyze the dataset and predict the outcome, and developed a decision support tool (DST) which is seamlessly integrated in the EHR DIPS Arena®. We intend to use the tool to inform the SDM between surgeons and patients about the indication for surgery (yes or no), to increase the proportion with a successful outcome.

The aim of the studies described in this clinical investigation plan (CIP) is to assess the safety and feasibility of the DST for use in a subsequent main effectiveness study.

The device
The DST (the device) is an integrate compound of software-solutions. Baseline data are registered by patients and surgeons on questionnaires integrated in DIPS Arena®, and transferred to NORspine. The data are also transferred (de-identified) to the AI-enabled prediction algorithm which operates in a cloud-based model hosting service. The algorithm has been trained and validated on a dataset from NORspine. The area under the curve for prediction of the main outcome (Oswestry disability index after12 months) in receiver operating characteristic analysis is very high (0.85) for LDH and moderate (0.72) for LSS. The model host also calculates outcomes (proportions with substantial, slight, or no improvement, and worsening) for the 50 cases with baseline variables most similar to the present case ("patients-like-me"). Finally, the individual prediction and the outcomes for the "patients-like-me" are transferred back and displayed in the regular user interface of DIPS Arena® for use in the SDM.

Clinical investigations
This CIP describes a feasibility study and clinical pilot (proof of concept) study. Both use convergent qualitative and quantitative mixed methods. The comparator is decision making in routine clinical practice, without use of the DST.

The feasibility study will include 20 patients with magnetic resonance imaging confirmed LDH or LSS referred for evaluation of the indication for surgery, and six surgeons who do the evaluations. The study will iteratively redesign the user interface of the DST until it is considered safe and feasible for use in the pilot study.

The pilot study will include 100 patients and nine surgeons at two study sites and evaluate whether the DST is safe and feasible for use in a possible subsequent main effectiveness study. It will also enable sample-size calculation for the main study.

Monitoring and adverse events
Possible foreseeable adverse events (AEs) are use of the DST on patients not fulfilling the inclusion criteria, erroneous or incomplete recording of baseline data, error in data transfer between the components, and misinterpretation of the prediction. Such AEs could misinform the SDM and cause unjustified recommendations about undergoing surgery or not.

The CIP describes data management and protection, the informed consent process, measures taken to avoid AEs, and monitoring and reporting in detail.

Registration and publication
This clinical investigation will be registered on ClinicalTrials.gov before the start of recruitment, and results will be disseminated by publication in international peer-reviewed open-access journals.

Norsk
Bakgrunn
En tredel av pasienter som blir operert for prolaps eller spinal stenose i ryggen oppnår ingen vesentlig bedring. Studier indikerer at godt informert samvalg kan forbedre utvelgelsen til kirurgisk behandling, og dermed resultatene. Mange har derfor utviklet prediksjonsalgoritmer, og noen av disse bruker kunstig intelligens (KI). De fleste er trent på små datasett, få har tilstrekkelig nøyaktighet, alle er til bruk i frittstående systemer eller nettbaserte applikasjoner og ingen er integrert i den elektroniske pasientjournalen (EPJ) eller implementert i rutinemessig klinisk praksis.

Nasjonalt kvalitetsregister for ryggkirurgi (NKR, eng. NORspine) har registrert en kohort på mer enn 69 000 ryggoperasjoner. Vi har brukt KI til å analysere datasettet og predikere operasjonsresultatet, og utviklet et beslutningsstøtteverktøy som er sømløst integrert i EPJ-systemet DIPS Arena®. Intensjonen er å bruke verktøyet til et bedre informert samvalg mellom kirurger og pasienter i vurderingen av operasjonsindikasjon (ja eller nei), for å øke andelen som får et godt operasjonsresultat.

Målet med studiene som beskrives i denne forskningsprotokollen (eng. Clinical investigation plan; CIP) er å undersøke om beslutningsstøtteverktøyet er trygt og anvendelig for bruk i en eventuell etterfølgende nasjonal effektstudie.

Utstyret
Beslutningsstøtteverktøyet (utstyret) er en sammensetning av mykvare-løsninger. Baselinedata blir registrert av pasienter og kirurger i digitale spørreskjemaer som er integrert i DIPS Arena®, og overført til NKR. Dataene blir også overført (avidentifisert) til en KI-basert prediksjonsalgoritme som driftes i en sky-basert modellvertstjeneste. Algoritmen er trent og validert på et datasett fra NKR. Arealet under kurven for prediksjon av hovedutfallsmålet (Oswestry disability index etter 12 måneder) i ROC-analyser er svært høyt (0,85) for prolaps og moderat (0,72) for spinal stenose. Modellvertstjenesten beregner også utfall (andelene med stor bedring, litt bedring, ingen tydelig bedring og forverring) blant de 50 pasientene med baseline-variabler mest lik den aktuelle pasienten («pasienter-som-meg»). Til slutt blir den individuelle prediksjonen og utfallet for «pasienter-som-meg» overført tilbake til DIPS Arena® og vist som et skjermbilde i det vanlige brukergrensesnittet slik at resultatet kan brukes i samvalget.

Kliniske studier
Denne protokollen beskriver en gjennomførbarhetsstudie og en klinisk pilotstudie. Begge bruker integrert kvalitativ og kvantitativ metode. Vi skal sammenligne beslutningsprosesser gjennomført med bruk av beslutningsstøtteverktøyet med rutinemessig klinisk praksis.

Gjennomførbarhetsstudien skal inkludere 20 pasienter som er henvist til vurdering med tanke på ryggkirurgi etter å ha fått påvist prolaps eller spinal stenose med magnettomografi-undersøkelse (MR), og seks kirurger som gjør slike vurderinger. I denne studien skal vi gjennomføre gjentatte forbedringer av verktøyets brukergrensesnitt til det vurderes som anvendelig i pilotstudien.

Pilotstudien skal inkludere 100 pasienter og ni kirurger på to sykehus, og evaluere om verktøyet er trygt og anvendelig for utprøving i en eventuell etterfølgende effektstudie. Den vil også fremskaffe foreløpige data om utfall til styrkeberegning for effektstudien.

Monitorering og uønskede hendelser

Mulige uønskede hendelser er bruk av verktøyet på pasienter som ikke fyller inklusjonskriteriene, feilaktig eller mangelfull registrering av preoperative data, feil i dataoverføring mellom komponentene, og feiltolking av resultatet av prediksjonen. Slike feil kan gi et feilaktig beslutningsgrunnlag for samvalget, og medføre ubegrunnede anbefalinger om å gjennomgå eller avstå fra kirurgisk behandling.

Protokollen beskriver databehandling og beskyttelse av personopplysninger, samtykkeprosessen, tiltak for å unngå uønskede hendelser, og monitorering og rapportering av studien i detalj.

Registrering og publisering

Studiene vil bli registrert på ClinicalTrials.gov før vi starter inklusjon av pasienter, og resultatene vil bli publisert i internasjonale fagfelle-vurderte åpen-tilgang tidsskrifter.

2. Identification and description of the investigational device

Background

Low back pain (with or without leg pain) caused by degenerative spondylosis is a major cause of non-fatal health loss [1]. In Norway, it is the most common cause for short-term sick leave and the second most common for disability benefits [2].

A Cochrane review found conflicting evidence on the effectiveness of surgical treatment [3], and only two thirds of the cases registered in the Norwegian registry for spine surgery (NORspine) report complete recovery or substantial improvement 12 months after the operation [4]. Lack of evidence-based guidelines and the fact that spine surgery is preference-sensitive contribute to variation in surgical rates, techniques and patients' outcomes between hospitals, health regions and countries [5-7].

Studies show that a complex interplay between patient- and treatment-related factors influence the outcome [8]. It is believed that well informed shared decision-making (SDM) can improve selection to surgery, and thus the outcomes, but this requires accurate methods for outcome prediction [9, 10]. Numerous algorithms for outcome prediction in spine surgery have therefore been developed [9, 11-14], and some are AI-based [15-18]. Most are trained on small data sets, and few have reached a satisfactory accuracy with an area under the curve (AUC) >0.80 in receiver operating characteristic (ROC) analysis. A recent review from our group shows that all are stand-alone or web-based systems not integrated in the electronic health record (EHR), and none are implemented in routine clinical practice [19].

NORspine is a large national clinical quality registry which covers all public and private providers and has a capture rate of 80 % [20]. The cohort comprises more than 69,000 cases (2023) and the dataset contains a comprehensive set of baseline-, process- and outcome variables recommended by the International Consortium for Health Outcome Measurement (ICHOM). The primary outcome is the Oswestry disability index (ODI), which is a disease-specific patient reported outcome measure (PROM) of functional disability [20]. NORspine thus provides a unique dataset for development of accurate AI-based outcome prediction.

We hypothesize that improved selection of patients who are likely to benefit from an operation will reduce the proportion of unsuccessful operations, and thus improve the outcomes after spine surgery. Recently, we explored use of the dataset in NORspine for AI-based prediction of success (complete recovery or substantial improvement) after operation for lumbar disk herniation, and achieved an AUC of 0.82 [21]. In parallel, we have developed a clinical decision-support tool (DST) for SDM which is seamlessly integrated in the regular user interface of the EHR-system DIPS Arena®.

Aims
The overreaching aim of our innovation and research is to complete the development of and assess the safety, feasibility and effectiveness of an AI-enabled DST for lumbar disc herniation (LDH) and lumbar spinal stenosis (LSS).

Function
The DST is an AI-enabled algorithm which is trained on data in NORspine to predict the outcome of surgery for LDH and LSS. The tool is seamlessly integrated in the regular user interface of DIPS Arena®. We intend to use the tool to inform and support SDM about the indication for surgery (yes or no) between spine-surgeons and patients during outpatient clinic consultations.

Details of development
The development was initiated by NORspine, and organized as a project by the Norwegian center for clinical artificial intelligence (SPKI). NORspine and SPKI are hosted by the University hospital of North Norway (UNN). The project partnered with all other stakeholders at the UNN, the Machine-learning group at UiT the Arctic university of Norway, the Northern Norway regional health authority ICT trust (HN IKT) and clinical IT-implementation project organization (HN FRESK), the Research department at DIPS ASA and the health-data analyst company Deepinsight AS.

We organized the development in six work-packages (WPs):

The prediction algorithm
Preliminary modelling verified that satisfactory accuracy (defined as an AUC > 0.80) for identifying successful outcomes (yes/no) can be achieved [21].

In WP1, we used a dataset of all recorded cases in NORspine operated from 01.01.2007 through 10.06.2022 (n=69,672) to develop the prediction algorithm. We excluded cases with obviously erroneous values for age, body weight and height in logical checks (n=141) and cases with missing data for the primary outcome (ODI) at baseline (n=877) or at both 3- and 12-month follow-up (n=12,486). For cases who responded after 3, but not 12 months, we imputed the ODI reported at 3-months as a substitute for the missing data at 12 months (the last value carried forward-method). We excluded cases not operated for LDH or LSS (n=8,455) and cases who underwent an emergency operation for cauda equina syndrome (n=459).

The dataset thus consisted of 47,254 cases operated for LDH (n=22,597) and LSS (n=24,657). We split the data in training/validation- (80%) and test-sets (20 %).

During the algorithm development, numerous models were tested for both dichotomous outcomes (treatment success or non-success) and prediction of a continuous outcome (ODI) 12 months after the operation. The most accurate results were achieved with an

extreme gradient boosting (XGBoost) model for prediction of the continuous outcome. The model for LDH reaches a very high accuracy with AUC of 0.85, root mean squared error (RMSE) 14 and R^2 0.27, while the model for LSS reaches a moderate accuracy (AUC 0.72, RMSE 15 and R^2 0.29).

The algorithm will not be changed during the feasibility and pilot study, but it can be re-trained for future versions, to utilize the continuously increasing amount of data in the registry to improve the accuracy. The AI-based prediction will be used to inform the SDM about possible outcomes on the individual level.

WP1 has also developed functionality for display of the outcomes for the 50 cases with baseline variables most similar to the present case (patients-like-me). Presentation of their outcomes will be used to inform the SDM about possible outcomes at the group level. The risk for indirect identification is mitigated by data minimization (e. g. collapsing continuous age into 5-year age categories) before data transfer to the model host, and anonymization is ensured by display only of proportions in the different outcome categories in DIPS Arena®.

Software integrations
WP2 digitalised NORspine's questionnaires and integrated them in DIPS Arena®. WP3 developed data transfer between DIPS Arena® and a cloud-based model hosting service for the algorithm, which operates within the Norwegian health network. WP4 will complete a design for presentation of the predictions in the user interface of DIPS Arena®. WP5 developed the necessary adaptions in NORspine for transfer of data from DIPS Arena®. It also developed application programming interfaces (API) for integrations between DIPS Arena® and NORspine, and between DIPS Arena® and the model hosting service.

WP6 resolved legal issues and clarified that the design can be used for clinical research, provided the necessary approvals are granted.

Figure 1 illustrates the design of the solution.

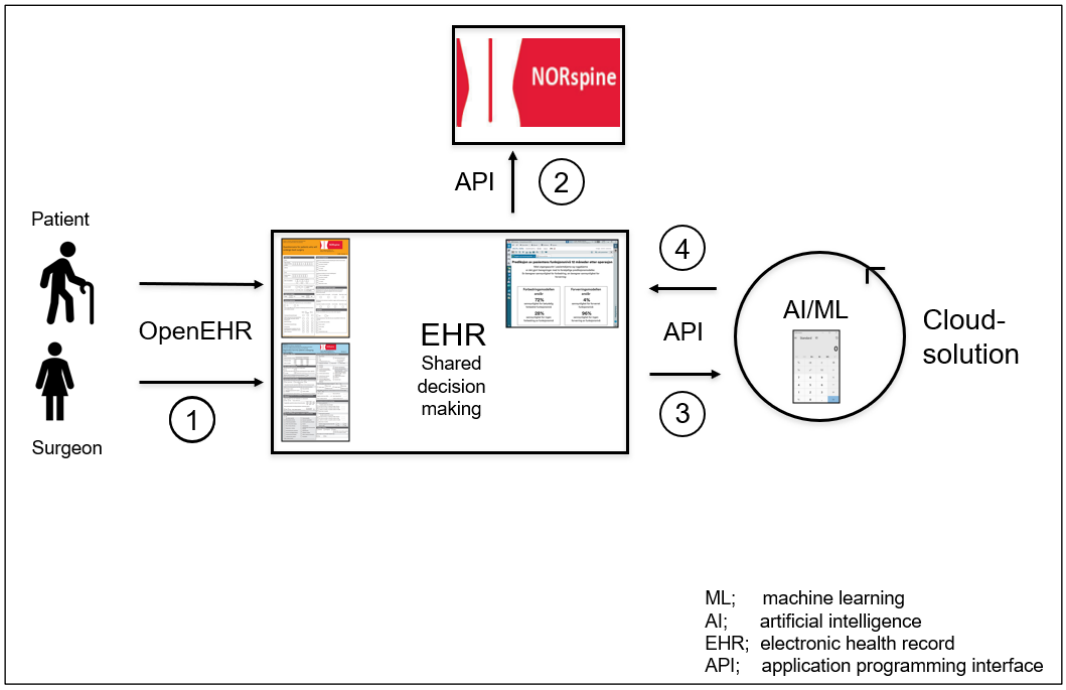

Figure 1. Design of an AI-enabled clinical decision support tool for spine surgery integrated in the EHR. Baseline data are registered by patients (online) and surgeons on questionnaires integrated in DIPS Arena® (1), and automatically transferred to NORspine (2). The data are also transferred (anonymized) to a prediction algorithm trained on the dataset in NORspine which operates in cloud-based model hosting service within the Norwegian health network (3). Finally, the prediction is transferred back to DIPS Arena® and displayed in the user interface (4).

Identification of the device

The device is version 1.0 of an integrate compound of software-solutions which has not received a commercial name yet. It is for the conduct of this clinical investigation entitled *Artificial intelligence (AI) enabled decision support tool for selection of patients for lumbar spine surgery* (short: *DST for spine surgery*; Norwegian: *Beslutningsstøtte ryggkirurgi*). Version logging and traceability will be according to DIPS ASA's regular established routine for software updates in DIPS Arena®.

Use

The surgeons who participate in the studies will be made aware that the final decision about the indication for surgery is the responsibility of the surgeon. The intention is to inform the SDM and retain a trustful surgeon-patient relation, and not to establish computer-paternalism [22].

The user of the DST must be a certified medical doctor and a specialist in neurosurgery or orthopedic surgery, or in training in one of these specialties. We will provide the surgeons a training session which explains the intention, limitations and possible risks of the tool. The limitations, e.g. that the DST is developed for use during outpatient clinic consultations, and not intended for use in emergency cases or in cases with a baseline

ODI ≤ 22 (LDH) or ≤ 14 (LSS) will be emphasized. The training session contains a demonstration of the use and provides hands-on training on a fictive case.

3. Justification for the design of the clinical investigation

The studies will use convergent mixed methods of qualitative and quantitative design.

Feasibility study (study 1)
The first study uses an iterative process to assess the safety and develop the feasibility of the DST for a subsequent clinical pilot study.

Integration of AI-based tools in healthcare providers' regular information technology (IT) infrastructure and clinical workflow with acceptable user-friendliness is prerequisite for adoption [23, 24], and user-interaction is important to assure acceptable usability and safety [25]. Knowledge about the feasibility is therefore necessary to determine whether evaluation of a tool in comprehensive clinical studies is justified [26]. The feasibility-study design is thus appropriate as the first introduction of a new device into clinical testing after lab-simulated development, because it evaluates the initial use of the device, while maintaining low risk for patients. The iterative design ensures that relevant feedback from surgeons, patients and other staff involved in the workflow at the outpatient clinic, leads to improvement of the DST, its usability and the related workflow.

Pilot study (study 2)
The second study is a clinical pilot (proof of concept) study.

Evaluation of a DST as safe and feasible in a one-center setting with a limited number of participating surgeons and patients is not sufficient to justify a large effectiveness-study, since the feasibility can vary between hospitals with different staffs and workflows. Further, estimates of effect sizes are necessary to evaluate whether a subsequent confirmatory main study is justified, and to enable sample size calculation [27].

The pilot study-design is appropriate to achieve this. After the feasibility study, a final prototype of the DST and the related workflow will be completed. The clinical pilot study design is appropriate for testing whether this prototype is safe and feasible for scaling to other hospitals and use among more surgeons.

4. Risks and clinical benefits of the investigational device and clinical investigation

Clinical benefits
Only 61 % of the cases registered in NORspine in 2021 reported complete recovery or substantial improvement 12 months after the operation. Among the remaining 39 %,

corresponding to about 2,000 patients per year (2022), 34 % reported little or no benefit, while a subgroup of 5 % reported worsening. Complications such as dural tear or wound infection occurred in 1-2 % and 5 %, respectively [4]. More serious complications, such as nerve root injury, transfusion- or reoperation-requiring bleeding, venous thromboembolism, respiratory or cardiac complications are rare (<1 %), and peri-operative death is extremely rare (<0,01 %).

In addition, there is substantial geographical variation both in the surgical rates, success rates and the occurrence of complications [4, 7].

These findings are in accordance with results from other large clinical quality registers for spine surgery.

Accordingly, improved selection of patients who are likely to benefit from an operation will reduce the high number of patients who undergo operations without achieving health gain, and prevent avoidable complications in this group. In addition, increased consistency in the selection can reduce unwarranted variation. Further, a reduced number of non-successful operations will free resources and treatment capacity, and potentially shorten waiting-times.

Adverse device effects
We do not expect serious adverse events (SAE) related to the DST since it is a non-invasive device (a software) providing information which is used to support decisions about whether a surgical intervention should be performed or not (EU's Medical device regulation (MDR) risk class IIb).

We will, however, register all possible adverse events (AE) in accordance with the MDR. We define an AE as an erroneous risk estimation which misinforms the SDM, and causes a risk for unjustified recommendations about the intervention. If not corrected by the surgeon's discretion, this can lead to unwarranted recommendations about undergoing an operation for patients who may not benefit, and contrary, to recommendations about not undergoing an operation for patients who might have benefited from such treatment.

Risks associated with participation in the clinical investigation
Data from NORspine and other large clinical quality registries show that surgeons have low accuracy in their selection of patients for surgery in current practice. Our preliminary analysis of the prediction model's accuracy indicates that the algorithm will perform significantly better than the individual surgeon's discretion [21]. We do not know how application of the prediction through the DST will influence the SDM and the final decisions about the indication for surgery. However, on group-level, we anticipate that participation in the study will improve the SDM and reduce the risk for undergoing unwarranted operations.

Nevertheless, for individual patients, a risk for unjustified recommendations remains.

Undergoing an unwarranted operation exposes the patient to the risk for complications explained above, and causes delay in the administration of other and potentially more efficient treatments.

Being advised against an operation that would have been efficient exposes the patient to a risk for prolonged duration of pain and disability, and worsening of neurological deficits such as muscle weakness, which carries a small risk for permanent disability. Prolonged sick-leave and more disability increases the risk for long term inability to work [28]. Assessment of these risks must consider that the main benefit of surgical versus conservative treatment is shortening of the duration of symptoms. After 1-3 years, there is no difference in outcomes between patients with LDH who undergo surgery and those who do not [29]. For patients with LSS, the natural history is less consistent [30].

Patients who are advised against an operation will be informed that they have access to re-assessment if the symptoms worsen or persist.

Possible interactions with concomitant medical treatments as considered under the risk analysis
The device does not interact with other medical treatments or devices.

Steps that will be taken to control or mitigate the risks
User training for the surgeons will mitigate the risks that surgeons and patients trust the DST uncritically. In the feasibility study, an investigator will observe the consultations and interview both surgeons and patients to understand how the SDM is perceived and used. The tool and the related workflow will be improved iteratively when shortcomings are identified.

Rationale for benefit-risk ratio
Current practice exposes patients for a considerable risk (39 %) of undergoing an operation without achieving health gain. We consider it likely that participation will reduce this risk. Contrary, participation can increase the risk for being advised against an operation that would have been efficient. The size of this risk in current practice is probably considerable, but unknown because NORspine and other prospective registers do not track patients who are evaluated for surgery, but not operated.

We consider the mitigating steps are adequate to reduce the risks to an acceptable level, and emphasize that patients who are advised against an operation have continuous access to re-assessment if the symptoms persist or worsen over time. Altogether, we consider the risk as very low and probably lower than undergoing assessment within regular current practice, due to the stringent systematic approach to assessment embedded in the study protocols.

5. Design of the clinical investigation

Feasibility study (study 1)
Artificial intelligence (AI) enabled decision support tool for selection of patients for lumbar spine surgery - a mixed methods feasibility study

Objectives
To evaluate and iteratively redesign

1. An AI-enabled DST for lumbar spine surgery
2. The related workflow in a spine surgery outpatient clinic

until the tool and the workflow is considered safe and feasible for use in a pilot prospective observational clinical (proof of concept) study.

Study design
This is a feasibility study using convergent interventional mixed methods qualitative and quantitative design with embedding to iteratively assess and improve the DST and the related workflow.

Primary outcomes
The primary outcomes (assessed with qualitative methods) are:

1. Surgeons' acceptability of the DST for a clinical pilot study (yes/no)
2. Patients' acceptability of the DST for a clinical pilot study (yes/no)

Timeframe
Acceptability will be assessed continuously, but finally evaluated towards the end of the study, after iterative redesign of the DST and the related workflow according to requirements identified with qualitative methods.

Secondary outcomes
The secondary outcomes (assessed with quantitative methods) are:

1. Surgeons' compliance rate
   (the proportion of consultations in which the surgeon uses the DST as intended)
2. Patients' compliance rate
   (the proportion of patients who complete the online questionnaire with the required information before the outpatient clinic visit)
3. Duration of the consultation (minutes)

Timeframe
The rates and the duration will be calculated as averages for the study period, and towards the end of the study, after iterative redesign according to requirements identified with the qualitative methods.

Outcome assessment
The qualitative data will be analyzed continuously and discretionary to identify requirements for iterative redesign of the DST and the related workflow, and to inform our understanding of the context (workflow, roles, workshare between surgeons and administrative staff, surgeon-patient relation, stakeholders trust in the intervention). The data from both methods will be connected throughout the development in the iterative process. The results will be presented by joint display to illustrate the feasibility of the DST, as described by Fetters et al. [31].

Follow-up
In this study, we do not plan study-specific follow-up of patients or surgeons after the outpatient clinic consultation. Patients elected for surgery will be followed up according to the UNN's regular clinical routine for lumbar spine surgery, which includes monitoring of PROMs according to the NORspine-protocol. Patients not elected for surgery will be referred back to their general practitioner for regular follow-up in primary care.

Data collection

Qualitative data
The qualitative data will be collected from:

1. Semi-structured interviews with surgeons, focusing on how they trust the prediction of the outcome, how the prediction influences the decision about the indication for surgery, how they perceive the DST's and the related workflow's safety and usability, and how the use influences time management during the consultation.
2. Semi-structured interviews with patients, focusing on the online form's usability, how they perceive and trust the DST, the SDM and the decision.
3. Investigators' observation of outpatient consultations with and without use of the DST
4. Unstructured feedback from involved and non-involved surgeons and other staff during morning conferences, in the outpatient clinic and in meetings with the software developers.

The semi-structured interviews will evolve around 6 to 12 open questions prepared in advance [32]. The interview guide will not exclude further comments from surgeons or patients, or supplemental questions from the investigator.

Quantitative data
Quantitative data on compliance rates and the duration of consultations will be collected by the investigators through participative observation. We will also record whether the

surgeon agrees with the prediction or not, and whether the outcome of the SDM (operation yes/no) agrees with the prediction of outcome categories.

The iterative process
User feedback will be given in meetings between the investigators and the software developers, and if necessary, the surgeons and the leadership of the outpatient clinic. We have not prespecified the frequency because it will depend on how the DST performs and integrates with the workflow. Resistance against changes in roles and the workflow is anticipated, and leadership involvement expected as necessary to overcome such problems.

Study site/location
University Hospital of North Norway, Tromsø, Norway, Sykehusvegen 38, 9019 Tromsø.

Subjects

Inclusion criteria

- Patients with MRI-confirmed LDH or LSS referred to UNN Tromsø for evaluation of the indication for surgery
- Specialists and physicians in training (for two years or more) in neurosurgery who evaluate such patients at the neurosurgical outpatient clinic

Exclusion criteria

- Patients unable to consent because of
    - Age < 18 years
    - Serious drug abuse of severe psychiatric disorders
    - Language barriers (patients who cannot speak or read Norwegian)
- Patients with a baseline ODI ≤14 (LDH) or ≤22 (LSS)
- Patients undergoing non-elective/emergency operations
- Patients with degenerative conditions other that LDH and LSS, fractures, primary infections, or malignant conditions of the spine
- Physicians in training with less than two years' experience with spine surgery

Recruitment and consent
We will recruit approximately 20 consecutive patients referred to the neurosurgical outpatient clinic. Eligible patients will be identified by the investigators by continuous screening of the list of patients who have been accepted for a consultation, but not allocated an appointment. Assessment of eligibility will be based on information in the referral.

Eligible patients will receive a postal letter with information about the study, and an invitation to participate and return a signed written consent in a pre-stamped envelope. Patients will be informed that not participating will not influence their treatment. They will be provided a direct telephone number for contact with a sub-investigator if they want additional information or want to ask questions. Patients who do not respond within one week, will be contacted per telephone by a sub-investigator to clarify whether they want to consent or not, before their appointment is scheduled. Those who consent will be informed electronically about the scheduling of their appointment, and invited to provide their baseline data as described below, under Procedures.

The six neurosurgeons will be recruited among the staff at the neurosurgical department by purposeful sampling done by the PI [33]. Initially, during the steepest learning curve of the iterative process, we will include two specialists and one physician in training who have participated in development of the DST and the study design. This follows the method of purposeful sampling, aiming for saturation or adequate information power [34]. When these surgeons begin to express satisfaction with the usability, we will recruit another two specialists and one physician in training who have not been previously involved. The surgeons must provide written informed consent before participation.

Completion of the clinical investigation
The study is complete when information saturation is reached, which we consider likely after inclusion of 20 patients and 6 surgeons. This is in accordance with recommendations of including a sample size of 6-12 to reach saturation in implementation research [35, 36]. This estimate can be influenced by the number and character of design- and workflow-iterations, implying that saturation can be reached earlier or later. The study will be continued until the DST and the workflow is considered feasible for the pilot study, or until it is considered not feasible for continued evaluation, based on assessments of the primary and secondary outcomes.

Time perspective
Scheduled study period: 01.12 – 31.04.2025.

Pilot study (study two)
Artificial intelligence (AI) enabled decision support tool for selection of patients for lumbar spine surgery - a mixed methods pilot study

Objectives

1. To evaluate the safety and feasibility of the DST, the related workflow and the study design for use in a subsequent main (effectiveness) study
2. To estimate the DST's effectiveness to enable sample-size calculations for a subsequent main study
3. To reach a decision about whether we should proceed with the main study or not

Study design
This is a prospective observational clinical pilot study using a mixed method design to evaluate whether the DST and the study design has the safety and feasibility needed to proceed with a subsequent larger multicenter main effect study, and to enable sample-size calculations for the main study. We increase the number of participating hospitals from one to two to gain experience with transferability of the DST, the workflow and the study design. The preliminary estimations of effectiveness will be done by comparing participant's outcomes 12 months after the operation with outcomes among cases undergoing routine preoperative evaluation, with use of NORspine as data source.

Primary outcomes
The primary outcomes (assessed with qualitative and quantitative methods) are:

1. Surgeons' acceptability of the DST and the related workflow for use in a larger multicenter main effect study (yes/no)
2. ODI raw- and change-scores 12 months after the operation compared between participants and non-participants

The sample-size calculation will be based on the ODI, and the evaluation of the study design on the surgeon's assessments of the acceptability. The final evaluation of whether the subsequent main effect study is justified (yes/no) will be based on an integral assessment of evaluation of the safety, the acceptability and the estimate of the sample-size required to show clinically relevant differences in effectiveness.

Secondary outcomes
The secondary outcomes will be assessed quantitatively 12 months after the operation:

1. Numeric rating scale (NRS) raw- and change-scores for back- and leg-pain
2. General perceived effect (GPE) scale score
3. Work status
4. EuroQoL five-dimension three-level questionnaire raw- and change-score

Outcome assessment
The qualitative data will be analyzed continuously and discretionary to assess the acceptability of the DST and the workflow after the surgeons have gained some experience with the tool. Results will be presented by joint display.

Follow-up
In this study, we do not plan study-specific follow-up of patients or surgeons after the outpatient clinic consultation. Patients elected for surgery will be followed up according to the hospitals' regular clinical routine for patients undergoing lumbar spine surgery,

while patients not elected for surgery will be referred back to their general practitioner for regular follow-up in primary care.

Patients who consent to registration in NOR-spine will be followed-up according to the registry's regular routine with questionnaires after 3 and 12 months. Patients who withdraw their consent to participation in the study will be excluded from the retrieval of outcome data from NORspine, and notified that they must withdraw their consent to registration in NORspine separately, if they want to do so.

Data collection

Qualitative data
The qualitative data will be collected from:

1. Semi-structured interviews with surgeons, focusing on how the surgeons trust the prediction of the outcome, how the prediction influences the decision about the indication for surgery, how they perceive the DST's and the related workflow's usability, and how the use influences time management.
2. Unstructured feedback from involved and non-involved surgeons and other staff during morning conferences and in the outpatient clinic.

The semi-structured interviews will evolve around 6 to 12 open questions prepared in advance [32]. The interview guide will not exclude further comments from surgeons or patients, or supplemental questions from the investigator.

Quantitative data
Use of the DST does not require that the patient consents to participating in NORspine because the algorithm allows calculation of the prediction based on the patient's baseline data, without data-exchange with NORspine. We assume, however, that most participants in this study will consent to registration in NORspine, since the capture rate was 80 % in 2022 [20]. Therefore, NORspine can be used as a data source for between groups comparisons of outcomes among participants versus non-participants. Patients report the ODI and all secondary outcomes in this study to NORspine, and we will apply for access to these data.

We will assess how much the patient and the surgeon perceived that the prediction influenced the SDM (on scales ranging from no to decisive influence). We will also record whether the surgeon agreed with the prediction (on a balanced Likert scale ranging from disagree completely to agree completely), and whether the surgeon considers that the decision (operation yes/no) would have been different (yes/no/uncertain) if the DST had not been used.

Study sites/locations

1. University Hospital of North Norway Tromsø, Sykehusvegen 38, 9019 Tromsø, Norway
2. Nordland Hospital Bodø, Parkvegen 95, 8092 Bodø, Norway

Subjects

Inclusion criteria

- Patients with MRI-confirmed LDH or LSS referred to UNN Tromsø or Nordland Hospital Bodø for evaluation of the indication for surgery
- Specialists and physicians in training (for two years or more) in neurosurgery or orthopedic surgery who evaluate such patients at the neurosurgical outpatient clinic at UNN Tromsø or the orthopedic outpatient clinic at Nordland Hospital Bodø

Exclusion criteria

- Patients unable to consent because of
    - Age < 18 years
    - Serious drug abuse of severe psychiatric disorders
    - Language barriers (patients who cannot speak or read Norwegian)
- Patients with a baseline ODI ≤14 (LDH) or ≤22 (LSS)
- Patients undergoing non-elective/emergency operations
- Patients with degenerative conditions other that LDH and LSS, fractures, primary infections, or malignant conditions of the spine
- Physicians in training with less than two years' experience with spine surgery

Recruitment and consent
This study will recruit approximately 100 consecutive patients referred to the participating hospitals. Considering the larger sample size and previous studies from NORspine, we do not expect skewed inclusion, and do not plan for strategic inclusion.

The planned enrollment is six neurosurgeons from UNN Tromsø and three orthopedic surgeons from Nordland Hospital Bodø. They will be recruited by the local PIs in Tromsø and Bodø, respectively. The neurosurgeons can overlap partially or completely with those participating in the feasibility study.

The recruitment will otherwise follow the same procedures as in the feasibility study.

Written informed consent will be obtained from patients and surgeons as in the feasibility study. Patients who consent to registration in NORspine also consent to participation in research, such as the current study. This allows us to collect outcome data from NORspine and link to the data collected specifically for this study. The consent also covers use of data from patients who are registered in NORspine, but not participating in the present study as controls.

Completion of the clinical investigation
The study will be complete when 100 patients and nine surgeons are included, or by 30. April 2026, whichever is reached first. The exact number is not important, but it should be large enough to enable sample size calculation for the subsequent main study, which by empirical estimates require 50-150 patients, depending on the effect sizes of interest.

Time perspective
Scheduled study period: 31.04.2025 – 30.06.2027

Measures to minimize bias

Feasibility study (study one)
This study uses a convergent interventional mixed methods qualitative and quantitative design, with weighting towards the qualitative methods, since assessment of the main outcomes is by semi-structured interviews.

Researcher bias in qualitative research include selection bias, confirmation bias, interpretation bias, reporting bias and dissemination bias [37].

Referrals from general practitioners (GPs) will be assessed by consultant surgeons according to routine practice. The list of patients who have been accepted for a consultation (but not allocated an appointment) will be screened continuously by the ph.d.-candidate, and eligible patients will be identified based on the information in the referrals. Eligible patients will be included consecutively to avoid selection bias. If the recruited patients show little variation in education, sex, age and ethnicity, we will consider strategic sampling among consecutively referred patients to increase representativity.

We use reflexivity (keeping a reflective diary and peer debriefing with a researcher not involved in the study) and triangulation to mitigate confirmation, interpretation and reporting bias. The triangulation method is use of multiple qualitative data sources (observation, interviews and feedback from e.g. morning conferences) and collection of supplemental quantitative data. Dissemination bias is mitigated by publishing the protocol on ClinicalTrials.gov and by following the publication policy outlined in chapter 15 in this CIP.

Pilot study (study two)
This study also uses convergent interventional mixed methods qualitative and quantitative design, but the design is more balanced since one main outcome (surgeons' acceptability) is assessed qualitatively, while the other (effectiveness) is assessed quantitatively by utilizing outcome data in NORspine.

The measures to mitigate minimize bias in the qualitative part are as in study one, except that recruitment will be strictly consecutive to avoid selection bias. We assume that the larger sample size (n=100) will ensure representativity. NORspine collects the outcome data independently of the researchers participating in the study. This mitigates reporting bias. A pilot study is explorative, and use of blinding or randomization is therefore not appropriate. Estimating effectiveness is not an aim, so steps to reduce bias of effect-estimates, such as randomization, serve no purpose. Patient who are operated during the same period at hospitals and registered in NORspine are controls. Propensity score matching will be done to ensure comparability. This method controls for hospital characteristics as a possible confounder.

Procedures
Table 1 (version 2, attached) describes the clinical procedures and workflow in the studies, including the differences between routine practice and the studies, in detail.

Routine clinical practice
In current routine practice, GPs refer patients with suspected LDH and LSS to MRI if they consider surgical treatment an option, based on the patient's history (including the severity and duration of symptoms) and the clinical findings. If MRI confirms one of the diagnoses, they refer the patient to a neurosurgical or orthopedic outpatient clinic for assessment of the indication for surgical treatment by a spine surgeon. Patients who are accepted for a consultation receive a short message service (SMS) notification about the scheduled appointment to their mobile phone.

At the outpatient clinic, a consultation which includes reviewing the patient's medical history, a physical examination and re-assessment of the MRI findings is done by the spine surgeon. The surgeon then evaluates the possible outcome after surgery by doing an integral discretionary assessment, and completes the SDM with the patient to reach a final conclusion about the indication for surgery (yes or no).

In this routine, patients who are selected for surgical treatment are invited to participate in NORspine when they are admitted for the operation. Patients and surgeons report baseline data to NORspine on paper questionnaires, but these data are not used in the pre-operative decision making.

Patients who consent to registration in NORspine will are invited to report patient reported outcome measures (PROMs) after 3 and 12 months by completing an electronic

questionnaire online. Patients unable to respond electronically are provided a paper questionnaire by regular mail.

Feasibility study (study one)
In this study, included patients will be invited to fill out an electronic version of NORspine's preoperative questionnaire online, before the consultation at the outpatient clinic. A SMS notification will include a link to NORspine's baseline questionnaire. Access requires logon to www.helsenorge.no. This is a pilot for a general service being developed by the National health network for collection of structured data from patients to the EHR.

At the beginning of the consultation, a summary of the information provided by the patient will be presented to the surgeon in the regular user-interface of DIPS Arena®, as a basis for the conversation. Next, the surgeon completes the preoperative part of NORspine's questionnaire for registration of medical information, and requests the prediction. Finally, the DST computes the outcome prediction and displays it in the user-interface of DIPS® Arena, and the information will be used to inform the SDM between the surgeon and the patient. Patients who are selected for surgery will receive written information about NORspine, and be invited to consent to participation, according to the established routine, before transfer of the data to the registry. After the consultation, the patient will be asked to participate in a semi-structured interview with the investigator.

Patients who consent to registration in NORspine will be invited to fill out the questionnaire again within the last two weeks before the operation, if the waiting time exceeds two weeks.

The subsequent process before undergoing surgery (or no surgery), the surgical procedures and the follow-up from NORspine does not deviate from routine clinical practice.

Pilot study (study two)
The procedure will be similar in the pilot study, except there is no interview with the patients after the consultation.

Compromising factors
A concern for the feasibility of both studies, is patients' ability to access and complete the questionnaire online before the consultation. Based on experience during the feasibility study, we will consider the need for assistance with completion of the questionnaire online or on paper (followed by punching of the data) prior to the consultation with the surgeon.

Investigational device and comparators
The AI-enabled DST integrated in DIPS Arena® is the only investigational device to be evaluated. We do not plan to change the input data or the ML algorithm during the study period. The user-interface of the DST and the related workflow will be iteratively redesigned during the feasibility study, and we will allow minor improvements also during the pilot study, if deemed necessary for transferability to the second study site.

The comparator is decision making in current routine clinical practice, without use of decision support tools.

Monitoring plan
The Research department at the UNN will be responsible for the monitoring on behalf of the sponsor. The sponsor has dedicated a monitor who will be responsible for overseeing the progress of the investigation and to verify that it is conducted, recorded, and reported in accordance with this CIP and subsequent amendments, written procedures NS-EN ISO 14155, and other applicable regulatory requirements. The monitoring will be conducted according to a written monitoring plan, which will be agreed when the study has been approved by the Norwegian medical products agency and the Ethics committee for clinical investigation of medicines and medical devices (REK KULMU).

6. Statistical considerations

Feasibility study (study one)
Descriptive statistics about the patients and surgeons, their acceptability of and compliance with the DST, the degree to which the surgeon agrees with the prediction or not, and the proportion in which the outcome of the SDM (operation yes/no) agrees with the prediction (success or failure) of the outcome, and the duration of the consultations will be calculated. Missing data will be reported. Imputation will not be done. Further analysis will not be required.

Pilot study (study two)
Analysis population: Participants who receive use of the DST in the SDM about the indication for surgery and subsequently undergo the operation (study group) will be compared to propensity score-matched cases registered in NORspine who were operated during the same time period without use of the DST (control group). Descriptive statistics of baseline data, the operation, complications and PROMs 3 and 12 months post-operatively will be calculated. Missing data will be reported, but imputation will not be done. The results will provide rough estimates of variability, effect size and correlation, which is necessary for sample size calculation for the subsequent main study. Significance will be defined to 0.05-level.

7. Data management and protection

The qualitative and quantitative data collected specifically for this study through interviews and participatory observation will be recorded electronically and stored de-identified in a project specific and password protected folder on a secure research server at the UNN, approved by the data protection officer (DPO). The identificatory key will be stored in a separate folder. Access is granted by the DPO and restricted to the investigators. The qualitative data will be saved in raw text format and the quantitative data in UNN's research data sampling system REDCap. Routines for this data management are well established, the complexity and amount of data is limited and there will be no data exchange with or export to other organizations.

In this project, the baseline data that are routinely collected on paper questionnaires from patients who consent to registration in NORspine, will instead be collected electronically and stored in DIPS Arena® at the UNN and the Nordland Hospital as regular EHR-content. This data storage adheres to the established regulatory requirements for EHRs. The UNN and the Nordland Hospital store their data separately and in accordance with DPIAs for their EHRs.

The data-transfer from DIPS Arena® to NORspine and between DIPS Arena® and the cloud-based model hosting service which operates the algorithm, uses APIs developed by HN IKT. This adheres to established regulatory requirements for data transfer between hospital enterprises and the national clinical quality registers hosted by HN IKT. The algorithm uses anonymized data. All patient identifiable data are thus stored in DIPS Arena® and NORspine.

Study participants who consent to registration in NORspine, will have their baseline data automatically copied and transferred from DIPS Arena® to NORspine.

In the pilot study (study 2), baseline- and outcome-data will be retrieved from NORspine, stored on UNN's secure research server as described above, and linked with the data collected specifically for the studies by use of the identificatory key. All analyses will then be done on de-identified data within the same secure environment as described above.

NORspine is hosted by the UNN and the CEO is the formal data controller. NORspine's data management is in accordance with a DPIA for the registry approved by the DPO. HN IKT is data processor on behalf of the UNN, according to a data processing agreement.

8. Amendments to the CIP

The Norwegian medical products agency and the REC shall be notified of all proposed changes to the approved CIP that are likely to have a substantial impact on the safety, health or rights of the study participants or on the robustness or reliability of the clinical data generated by the investigation, as required in Article 75 of the MDR. The coordinating investigator must wait for 38 days before any modification is implemented (unless any of the exceptions in Article 75 applies) or for approval of the modifications, whichever comes first, before implementing the changes. A CIP with approved substantial modifications will be filed as a new main version (e.g. CIP code change from 01 to 02).

Non-substantial amendments to the CIP can be suggested by the PI and approved by the sponsor. Changes cannot be implemented without approval from the sponsor. A non-substantial modification will be filed as a new sub-version (e.g. CIP code change from 01.0 to 01.1).

It is the coordinating investigator's responsibility to inform all PIs and sub-investigators about amendments to the CIP, and to oversee consistent implementation of approved changes.

9. Deviations from the CIP

The investigators are not allowed to deviate from the CIP except to protect the rights, safety, and well-being of human subjects under emergency circumstances, when the investigator may deviate without prior approval from the sponsor. Waivers from the CIP are not permitted.

Sub-investigators and PIs must report deviations from the CIP to the coordinating investigator within one day, and the coordinating investigator must notify the sponsor on the first subsequent work day. The PIs must record all such deviations in the hospitals' electronic reporting system (Docmap) as a non-medical adverse event (AE) within one work day after their occurrence (the two study sites use similar systems for recording and analyses of AEs). A simplified event analysis according to the hospitals' regular routines will be done within two weeks. The coordinating investigator is responsible for the analyses.

In the case of unwarranted deviations from the CIP, the coordinating investigator is responsible for notifying and correcting sub-investigators. Repeated unwarranted deviations may lead to disqualification of sub-investigators.

10. Device traceability and accountability

Use of the DST will be limited to the principal investigators, the clinical sub-investigators and the surgeons who participate in the study by user access control, i.e. access to the tool will only be available for specific investigators and surgeons logged on the hospitals' IT-system and the DIPS Arena® EHR (two steps) with their personal user ID and password. These users will be given access to open and register data in the documents stored in DIPS Arena® and access to the prediction presented by the DST.

The DST will be labeled *«Exclusively for clinical investigation»* in the user-interface. The sponsor provides the study sites with a written instruction for use (IFU) and technical support. The principal investigators shall keep records documenting names of the persons who have the DST available on their EHR-account, who use it during the study, dates of use, which version they use, and subject (patient) identification.

HN IKT will keep a log specifying which users have access at any time. They will also log lookups made by users of the documents.

11. Statements of compliance

This clinical investigation will be conducted in compliance with the MDR. It also complies with current national and international regulations governing clinical investigations, the ethical principles that have their origin in of the *Declaration of Helsinki* [38] and the standard ISO 14155:2020 *Clinical investigation of medical devices for human subjects – Good clinical practice*.

The clinical investigation will not begin until the required regulatory and ethical assessments have been completed with non-negative outcomes, in accordance with the MDR and national legislation. Any additional requirements imposed by the REK or a regulatory authority will be followed, if appropriate.

Insurance is covered according to the standard public insurance at the University Hospital of North Norway (Norwegian: Norsk pasientskadeerstatning).

12. Informed consent process

The principal investigators will ensure that the participants are given full and adequate oral and written information about the clinical investigation, its purpose, risks and benefits, as well as inclusion and exclusion criteria. Subjects will be informed that they are free to discontinue their participation at any time without having to provide a reason, and that this will have no consequence for their treatment (patients) or employment (surgeons).

Subjects will be given the opportunity to ask questions and allowed time to consider the provided information and participation. If the person chooses to participate, both the subject and the investigator shall sign the informed consent form. A copy of the information and a copy of the informed consent form will be provided to the subject. The subject's signed and dated informed consent must be obtained before performing any activity specific to the clinical investigation. The process will be documented in the subject's source documents and the signed informed consents will be maintained with the essential documents. If new information becomes available that can significantly affect a subject's future health and medical care, that information shall be provided to the affected subject(s) in written form. If new information is added to the clinical investigation, the subject has the right to reconsider whether he/she will continue their participation.

Participation or non-participation in the NORspine follows established routines independently of this study.

Informed consent process for vulnerable populations

Patients are generally considered vulnerable. According to the established inclusion- and exclusion criteria for NORspine, we will not include children <18 years or patients unable to consent because of serious drug abuse, severe psychiatric disorders or language barriers [20]. A special informed consent process for particularly vulnerable populations is thus not needed.

13. Adverse events, adverse device effects and device deficiencies

Definitions

Adverse event

An adverse event (AE) is any untoward medical occurrence, disease or injury or clinical signs, in subjects, users or other persons, in the context of a clinical investigation, whether or not related to the investigational device.

This definition includes events that are anticipated as well as unanticipated. This definition includes events occurring in the context of a clinical investigation related to the investigational device, the comparator or the procedures involved.

Adverse device effect

An adverse device effect (ADE) is any AE related to the use of an investigational medical device. This definition includes adverse events resulting from insufficient or inadequate instructions for use, installation, or any malfunction of the device. This definition includes any event resulting from use error or from intentional misuse.

In these studies, ADEs are any AE related to the DST and use or misuse of it, causing an erroneous outcome prediction which misinforms the SDM, and causes a risk for unjustified recommendations about the intervention (undergoing or not undergoing surgery). If not corrected by the surgeon's discretion, this can lead to unwarranted advice recommending an operation for patients who will not benefit, or advice against an operation for patients who would have benefited from such treatment.

Identifying such AEs is difficult, because for individuals, we will not know the outcome of the alternative treatment strategy, since undergoing one type of treatment (surgical or conservative) precludes observation of the outcome of the other. The crossover-rate from conservative to surgical treatment can be indicative, and will be registered, but assessing it will be difficult, since spondylosis is dynamic, implying that the indication for surgery change over time for the same individual.

Recommendations considered to be obviously in conflict with present evidence and established best practice will be identified qualitatively through the participatory observation and interviews with the surgeons, counted and reported as possible AEs.

Serious adverse event
A serious AE (SAE) is any AE that led to any of the following:

a) Death
b) Serious deterioration in the health of the subject, that resulted in any of the following:
    i. Life-threatening illness or injury
    ii. Permanent impairment of a body structure or a body function
    iii. Hospitalization or prolongation of patient hospitalization
    iv. Medical or surgical intervention to prevent life-threatening illness or injury or permanent impairment to a body structure or a body function
    v. Chronic disease
c) Fetal distress, fetal death or a congenital physical or mental impairment or birth defect

Serious adverse device effect
A serious adverse device effect (SADE) is an ADE that has resulted in any of the consequences characteristic of a SAE. Such events related to procedures imposed by the clinical investigation plan but not with the use of the device shall not be considered a SADE. Accordingly, SAEs occurring as complications to the operations are not considered SADEs, and thus not reported as such.

Unanticipated Serious Adverse Device Effect
An unanticipated SADE is an effect which by its nature, incidence, severity, or outcome has not been identified in the current risk assessment. SAEs related to procedures

imposed by the clinical investigation plan but not with the use of the DST shall not be considered SADEs.

Device Deficiency
A device deficiency (DD) is any inadequacy in the identity, quality, durability, reliability, safety or performance of an investigational device, including malfunction, use errors or inadequacy in information supplied by the manufacturer.

In the present study, any obviously erroneous prediction produced by the DST, or erroneous predictions produced by user-error from patients or surgeons, should be recorded and reported accordingly.

Recording and Reporting

Recording
All investigators will record all AEs, SAEs, SADEs and DDs according to the definitions above in the hospitals' electronic reporting system for AEs (Docmap) within one work day after their occurrence. A simplified event analysis according to the hospitals' regular routines will be done within two weeks. The two study sites use similar systems for recording and analyses of AEs. The coordinating investigator is responsible for the recording and the analyses.

Reporting
The principal investigators will report all SAEs, SADEs and DDs to the sponsor (UNN's research director) immediately, but not later than three calendar days after the investigation site's study personnel's awareness of the event.

The sponsor will report to the Norwegian medical products agency:

- Any SAE and SADEs that has a causal relationship with the DST or the investigation procedure, or where such causal relationship is reasonably possible
- Any DD that might have led to a SAE if appropriate action had not been taken, intervention had not occurred, or circumstances had been less fortunate
- Any new findings in relation to any event referred to above

Reporting by the sponsor will be done by the *Summary reporting form* (MDCG 2020-10/2). The form will be updated for each reportable event and for new findings/updates to already reported events. For events that indicate an imminent risk of death, serious injury, or serious illness and that requires prompt remedial action for other patients/subjects, users or other persons or a new finding to it will be reported immediately, but not later than two calendar days after awareness by the sponsor of a new reportable event or of new information in relation with an already reported event. Any other reportable events or a new finding/update to it will be reported immediately, but not later than seven

calendar days following the date of awareness by the sponsor of the new reportable event or of new information in relation with an already reported event.

Assessment of causality

The relationship between each AE and the investigational device, the comparator and the investigation procedure will be assessed and recorded by the coordinating investigator and sponsor. The sponsor and investigator will distinguish between SAEs related to the DST and those related to the procedures (relatedness to both is possible).

Each SAE will be classified according to four different levels of causality:

1. Not related

   Relationship to the DST, comparator or procedures can be excluded when:

   i. The event has no temporal relationship with use of the DST, or the procedures related to use of it
   ii. The SAE does not follow a known response to use of the DST (if the response pattern is previously known) and is biologically implausible
   iii. The SAE can be attributed to another cause (e.g. an underlying or concurrent illness/ clinical condition, an effect of another device, drug, treatment or other risk factors)
   iv. The event does not depend on a prediction given by the DST

   In order to establish the non-relatedness, not all the criteria listed above must be met at the same time.

2. Possible

   The relationship with the use of the DST or the related procedures is weak, but cannot be ruled out completely. Alternative causes are also possible (e.g. an underlying or concurrent illness, clinical condition or/and an effect of another device, drug or treatment). Cases where relatedness cannot be assessed, or no information has been obtained shall also be classified as possible.

3. Probable

   The relationship with the use of the DST or the related procedures or the comparator, seems relevant and/or the event cannot be reasonably explained by another cause.

4. Causal relationship

   The SAE is associated with use of the DST or the related procedure, or the comparator or with procedures beyond reasonable doubt when:

   i. The event is a known side effect of the DST or the related procedures
   ii. The event has a temporal relationship with its use
   iii. The SAE follows a known response pattern to use of the DST (if the pattern is previously known)
   iv. Other possible causes (e.g. an underlying or concurrent illness, clinical condition or/and an effect of another device, drug or treatment) have been adequately ruled out
   v. Harm to the subject is due to error in use or interpretation of the prediction produced by the DST
   vi. The event depends on an erroneous prediction given by the DST

In order to establish the relatedness, not all the criteria listed above might be met at the same time.

List of foreseeable Adverse events
In these studies, device-specific adverse effects are defined as erroneous outcome predictions which misinforms the SDM, and causes unjustified recommendations about the intervention (undergoing or not undergoing surgery).

Possible foreseeable AEs are:

1. User-error from a patient or a surgeon causes erroneous or incomplete recording of input data to the prediction algorithm, and this leads an erroneous prediction of the outcome
2. Errors in data transfer between the different components of the SDM causes erroneous or incomplete input data to the prediction algorithm, and this leads to an erroneous prediction of the outcome
3. The surgeon and/or the patient misinterpret the output from the prediction, and base the SDM on this misinterpretation
4. The DST is applied to patients not fulfilling the inclusion criteria or to patients who should have been excluded. This would cause an irrelevant outcome prediction, and misinform the SDM.

14. Premature termination of the clinical investigation

The coordinating investigator, the sponsor, the Norwegian medical products agency, and the REC can all suspend or prematurely terminate or halt the investigation, if deemed necessary. The monitor can advise these actors to terminate the study. Premature termination of the investigation must be for significant and documented reasons.

If suspicion of an unacceptable risk to subjects arises, or when so instructed by the Norwegian medical products agency, the sponsor will suspend the investigation while the risk is assessed. The sponsor will terminate the investigation if an unacceptable risk is confirmed, and inform all investigators.

The sponsor shall consider terminating or suspending the participation of a particular investigation site or investigator if the monitoring identifies serious or repeated deviations on the part of a site or an investigator. If the suspension or premature termination was in the interest of safety, the sponsor shall inform all other principal investigators.

If, in the opinion of the principal investigator, clinical observations suggest that it may be unsafe to continue the investigation at the site, the investigator may terminate the site's participation after consultation with the sponsor. A written statement documenting the reasons for such termination shall be provided to the sponsor.

If the investigation is prematurely terminated, the investigators shall promptly inform the subjects and take necessary steps to finalize their engagement in the investigation. All relevant investigation material must be collected, and accountability completed.

If the clinical investigation is interrupted or terminated prematurely, the sponsor will report to the Norwegian medical products agency within 15 days together with a justification. If the sponsor has temporarily halted or prematurely terminated the investigation on safety grounds, the Norwegian medical products agency will be informed within 24 hours. A clinical investigation report will be prepared within three months of the early termination or temporary halt, irrespective of the results. In the event that the investigation is restarted within three months after the temporary halt, the sponsor does not have to submit a clinical investigation report until the clinical investigation has been completed. The final clinical investigation report shall include details with respect to the temporary halt. If relevant, affected patients will be followed up after termination of the investigation by either telephone or outpatient clinic, depending on what the coordinating and principal investigators find appropriate.

We consider the risk of events necessitating a premature or temporary termination as low due to the low-risk nature of the DST. However, we cannot preclude that the DST will be evaluated significantly less useable by surgeons who have not been participated in the development than by the involved surgeons. In such a case, there is a risk that the iterative redesign process can identify technical obstacles that are difficult to overcome. This can lead to temporary halting of the feasibility study (study 1), and in a worst-case scenario to termination of the study.

15. Publication policy

This clinical investigation will be registered on ClinicalTrials.gov before the start of recruitment activities and the content will be updated throughout the conduct of the clinical investigation and the results entered at completion of the clinical investigation.

A clinical investigation report according to MDR article 77 will be filed to the Norwegian medical products agency by the sponsor within one year after closure of the pilot study (study 2) and the results will thus be publicly available.

The two studies will be submitted as separate publications to international peer-reviewed open-access journals within one year after completion of the recruitment (the feasibility study) and one year after completion of 12-month follow-up (the pilot study), independently of the findings. This means that negative outcomes will be published to mitigate publication bias. The sponsor will not take an active role in the publishing. The coordinating investigator is responsible for publishing the studies, and the criteria for authorship will follow the Vancouver Recommendations [39].

16. List of technical and functional features of the device

The DST is an integrate compound of software-solutions in the DIPS Arena® EHR, the national clinical quality register NORspine, and an AI-enabled prediction algorithm specifically developed for the DST.

Technical architecture
The DST depends on four software components (Figure 2):

1. DIPS Arena® which stores the questionnaires in a structured format and presents the user interface for the tool.
2. The user interface, which provides the users access to the questionnaires and the predictions.
3. The backend service, which retrieves predictive variables from the questionnaires, transfers them to the cloud-based model hosting service for the AI-enabled algorithm, and stores the predicted outcome in the patient record.
4. The model hosting service, which performs the prediction.

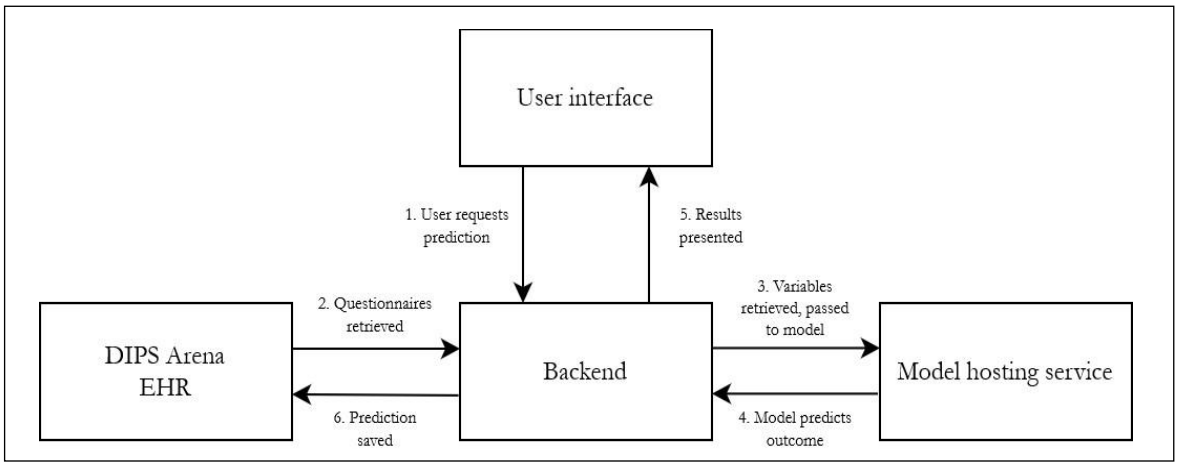

Figure 2. The system architecture. Major components and the flow of data between them.

Data flow

The software requires that both the patient and the surgeon complete the baseline questionnaires in DIPS Arena® before a prediction of the outcome can be requested:

1. The user interface is shown as an embedded application in DIPS Arena®, and the surgeon selects the questionnaires which he/she wants the prediction to be based on (by default, the last registered).
2. The backend service retrieves the questionnaires and extracts the variables required by the algorithm.
3. The backend transfers these variables (de-identified) to the model hosting service.
4. The model hosting service operates the algorithm, computes an outcome prediction and provides explanations for the prediction.
5. The backend service formats these results, transfers them to DIPS Arena® and presents them in the user interface, along with a summary of the information registered in the questionnaires.
6. Finally, the prediction is stored as a separate document in the patient's DIPS Arena® EHR.

Details of development

- The user interface is a web application developed in the Svelte framework, which implements the Substitutable medical applications and reusable technologies (SMART) on Fast healthcare interoperability resources (FHIR) standard for authentication and resource retrieval.
- The backend service is a Python application that manages the data flow described above. It hosts the user interface and, once the user is authenticated, provides it access to predictions for specified questionnaires. It retrieves questionnaires from DIPS Arena®, selects the appropriate variables from the questionnaires, and transfers them to the prediction algorithm.

- The model host is a Python application, which provides a representational state transfer (REST) interface to computing predictions from an extreme gradient boosting (XGBoost) model.
- DIPS Arena® manages authentication and authorization for the DST. Before use, a system administrator must authorize the user to access the application. When the user starts the DST, the tool requests an authorization token for the session from DIPS Arena®. Without this token, the user is not allowed to request patient questionnaires and cannot use the tool.
- The backend service and user interface are deployed in the same network infrastructure as DIPS Arena®, while the cloud-based model hosting service is accessible through the National health network.
- Data transfer between the components use an encrypted hypertext transfer protocol secure (HTTPS), through REST interfaces defined for each component with explicit API specifications.

17. Attachments

1. Table displaying the workflow and roles in current routine practice and in the clinical investigation

18. Bibliography

1. Tollånes, M.C., et al., *Disease burden in Norway in 2016.* Tidsskr Nor Laegeforen, 2018. **138**(15).
2. Brage, S., et al., *[Musculoskeletal disorders as causes of sick leave and disability benefits].* Tidsskr Nor Laegeforen, 2010. **130**(23): p. 2369-70.
3. Gibson, J.N., I.C. Grant, and G. Waddell, *The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis.* Spine (Phila Pa 1976), 1999. **24**(17): p. 1820-32.
4. Solberg, T.K., et al., *Årsrapport 2022: Nasjonalt kvalitetsregister for ryggkirurgi, resultater og forbedringstiltak.* Septentrio Reports, 2023(1).
5. Wennberg, J.E., *Unwarranted variations in healthcare delivery: implications for academic medical centres.* Bmj, 2002. **325**(7370): p. 961-4.
6. Weinstein, J.N., et al., *United States' trends and regional variations in lumbar spine surgery: 1992-2003.* Spine (Phila Pa 1976), 2006. **31**(23): p. 2707-14.
7. Ingebrigtsen, T., et al., *Treatment rates for lumbar spine surgery in Norway and Northern Norway Regional Health Authority 2014-18.* Tidsskr Nor Laegeforen, 2020. **140**(17).
8. Yap, Z.L., et al., *The role of the social determinants of health in outcomes of surgery for low back pain: a systematic review and narrative synthesis.* Spine J, 2022. **22**(5): p. 793-809.
9. McGirt, M.J., et al., *An analysis from the Quality Outcomes Database, Part 1. Disability, quality of life, and pain outcomes following lumbar spine surgery: predicting likely individual patient outcomes for shared decision-making.* J Neurosurg Spine, 2017. **27**(4): p. 357-369.
10. Stacey, D., et al., *Decision aids for people facing health treatment or screening decisions.* Cochrane Database Syst Rev, 2024. **1**(1): p. Cd001431.
11. Khor, S., et al., *Development and Validation of a Prediction Model for Pain and Functional Outcomes After Lumbar Spine Surgery.* JAMA Surg, 2018. **153**(7): p. 634-642.
12. Carreon, L.Y., et al., *Performance of the streamlined quality outcomes database web-based calculator: internal and external validation.* Spine J, 2023.
13. Lubelski, D., et al., *Patient-specific prediction model for clinical and quality-of-life outcomes after lumbar spine surgery.* J Neurosurg Spine, 2021. **34**(4): p. 580-588.
14. Fritzell, P., J. Mesterton, and O. Hagg, *Prediction of outcome after spinal surgery-using The Dialogue Support based on the Swedish national quality register.* Eur Spine J, 2022. **31**(4): p. 889-900.
15. Pedersen, C.F., et al., *PROPOSE. Development and validation of a prediction model for shared decision making for patients with lumbar spinal stenosis.* N Am Spine Soc J, 2024. **17**: p. 100309.
16. Pedersen, C.F., et al., *Applied Machine Learning for Spine Surgeons: Predicting Outcome for Patients Undergoing Treatment for Lumbar Disc Herniation Using PRO Data.* Global Spine J, 2022. **12**(5): p. 866-876.
17. De Silva, T., et al., *SpineCloud: image analytics for predictive modeling of spine surgery outcomes.* J Med Imaging (Bellingham), 2020. **7**(3): p. 031502.

18. Müller, D., et al., *Development of a machine-learning based model for predicting multidimensional outcome after surgery for degenerative disorders of the spine.* Eur Spine J, 2022. **31**(8): p. 2125-2136.
19. Toh, Z.A., et al., *Clinical Decision Support System Used in Spinal Disorders: Scoping Review.* J Med Internet Res, 2024. **26**: p. e53951.
20. Mikkelsen, E., et al., *The Norwegian registry for spine surgery (NORspine): cohort profile.* Eur Spine J, 2023. **32**(11): p. 3713-3730.
21. Berg, B., et al., *Machine Learning Models for Predicting Disability and Pain Following Lumbar Disc Herniation Surgery.* JAMA Netw Open, 2024. **7**(2): p. e2355024.
22. Čartolovni, A., A. Tomičić, and E. Lazić Mosler, *Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review.* Int J Med Inform, 2022. **161**: p. 104738.
23. Eiskjær, S., et al., *Usability and performance expectancy govern spine surgeons' use of a clinical decision support system for shared decision-making on the choice of treatment of common lumbar degenerative disorders.* Front Digit Health, 2023. **5**: p. 1225540.
24. Brøgger, H.M.B., et al., *[Not Available].* Tidsskr Nor Laegeforen, 2023. **143**(14).
25. Berkovsky, S. and E. Coiera, *Moving beyond algorithmic accuracy to improving user interaction with clinical AI.* PLOS Digit Health, 2023. **2**(3): p. e0000222.
26. Bowen, D.J., et al., *How we design feasibility studies.* Am J Prev Med, 2009. **36**(5): p. 452-7.
27. Arain, M., et al., *What is a pilot or feasibility study? A review of current practice and editorial policy.* BMC Med Res Methodol, 2010. **10**: p. 67.
28. Hara, S., et al., *Return to Work After Surgery for Cervical Radiculopathy: A Nationwide Registry-based Observational Study.* Spine, 2023. **48**(4).
29. Jacobs, W.C.H., et al., *Surgery versus conservative management of sciatica due to a lumbar herniated disc: a systematic review.* European Spine Journal, 2011. **20**(4): p. 513-522.
30. Wessberg, P. and K. Frennered, *Central lumbar spinal stenosis: natural history of non-surgical patients.* Eur Spine J, 2017. **26**(10): p. 2536-2542.
31. Fetters, M.D., L.A. Curry, and J.W. Creswell, *Achieving integration in mixed methods designs-principles and practices.* Health Serv Res, 2013. **48**(6 Pt 2): p. 2134-56.
32. Malterud, K., *[Qualitative methods in medical research--preconditions, potentials and limitations].* Tidsskr Nor Laegeforen, 2002. **122**(25): p. 2468-72.
33. Palinkas, L.A., et al., *Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research.* Adm Policy Ment Health, 2015. **42**(5): p. 533-44.
34. Malterud, K., V.D. Siersma, and A.D. Guassora, *Sample Size in Qualitative Interview Studies: Guided by Information Power.* Qual Health Res, 2016. **26**(13): p. 1753-1760.
35. Guest, G., A. Bunce, and L. Johnson, *How Many Interviews Are Enough?:An Experiment with Data Saturation and Variability.* Field Methods, 2006. **18**(1): p. 59-82.
36. Hamilton, A.B. and E.P. Finley, *Qualitative methods in implementation research: An introduction.* Psychiatry Res, 2019. **280**: p. 112516.
37. Johnson, J.L., D. Adkins, and S. Chauvin, *A Review of the Quality Indicators of Rigor in Qualitative Research.* Am J Pharm Educ, 2020. **84**(1): p. 7120.

38. *World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.* Jama, 2013. **310**(20): p. 2191-4.
39. ICMJE, *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. 2023: https://www.icmje.org.